NASAL CODA MERGER IN HANGZHOU MANDARIN: A DYNAMIC ULTRASOUND STUDY

Yin Huang[1], Liang Ma[2], Christopher Carignan[1]

[1]University College London, [2]Fudan University
yin.huang.21@ucl.ac.uk, liangma@fudan.edu.cn, c.carignan@ucl.ac.uk

ABSTRACT

Syllable-final nasals /n/ and /ŋ/ are reported to be neutralised in Mandarin varieties influenced by Wu dialects, among which is Hangzhou Mandarin. Previous articulatory research has shown supporting evidence for neutralisation, but only for single time points. In the current paper, we introduce a dynamic method of analysing ultrasound images, in order to determine whether neutralisation is present throughout the entire temporal interval of nasal coda production. This method quantitatively captures the movement of the tongue during the speech production process, which helps to examine the articulation distinction of different speakers. This study provides conclusive evidence that the place of articulation contrast between alveolar and velar coda nasals is neutralised in Hangzhou Mandarin.

Keywords: Ultrasound, nasal coda, merger, Mandarin varieties.

1. INTRODUCTION

Standard Mandarin allows alveolar nasal /n/ and velar nasal /ŋ/ in the syllable-final coda position. And the contrastive distribution of the /n/ and /ŋ/ carries the important function of distinguishing word meanings [7]. Standard Mandarin has 177 syllables with nasal codas, accounting for 44% of the total 405 regular syllables in Mandarin phonology. However, this contrast is observed to be reduced or neutralised in some regional varieties of Mandarin.

Wu dialects are spoken in southeastern China by more than 85 million people [20]. The phonological system of Wu dialects has vast differences from that of Standard Mandarin, one of which is that Wu dialects allow only a single nasal segment in codas [2,14]. This phoneme is usually transcribed as alveolar nasal /n/. Hangzhou is a major city in the Wu-speaking area, where the local dialect shares the common nasal coda limitation of Wu dialects [19]. Before the establishment of Standard Mandarin as a nationwide lingua franca, Wu dialects used to be the common language of this region. So Mandarin speakers in Wu-speaking areas show some difficulty distinguishing alveolar nasal /n/ and velar nasal /ŋ/, including Hangzhou Mandarin speakers.

From an acoustic perspective, the syllable-final nasals in Standard Mandarin are weakly produced with incomplete oral closure, meaning they have lower intensity than regular nasal consonants and more blurred vowel-nasal boundaries [4]. Thus, determining the place of articulation of the coda nasal by its acoustic pattern becomes rather complicated. As a result, most previous studies rely on perception methods to evaluate the neutralisation of nasal codas in accented Mandarin [9,12,17]. The results from these perceptual studies have suggested that in Mandarin varieties influenced by Wu dialects, the nasal coda /n/ and /ŋ/ are widely neutralised. The nasal coda merger is affected by some social factors, one of which is age, but the merger does exist in all age groups. The Wu dialect speakers tend to pronounce the velar nasal /ŋ/ as alveolar /n/, and their perceptual sensitivity to the /n/-/ŋ/ contrast is much lower than Standard Mandarin speakers.

The only articulation study of nasal coda merger in Wu-accented Mandarin is Faytak et al. [8]. They use a dimensionality reduction technique to process the midsagittal ultrasound data, comparing the tongue shape in the articulation of nasal coda produced by Shanghai Mandarin speakers and Standard Mandarin speakers. The results support the claim that the place contrast of nasal coda in Shanghai Mandarin is reduced in certain vowel environments. In this study, the midpoint of each nasal coda token was taken as a representative "magic moment" for determining tongue shape, thus disregarding possible dynamic factors involved in the neutralisation. However, given that these frames were chosen in an a priori manner, it is not certain whether the tongue shapes at these magic moments are indeed representative of the articulatory reality of the merger. This is due to the fact that, firstly, the vowel-nasal boundary of nasal coda syllables in Mandarin is hard to find, so the segmentation of nasal coda is of great difficulty. Secondly, the production movement of the coda nasal is too rapid for the tongue to achieve a stable position, meaning that a single midpoint frame may not be convincing enough to indicate the place of articulation of a specific nasal coda token.

Therefore, this paper attempts to propose a new approach of ultrasound data analysis that can dynamically track tongue movement to reflect the nasal coda's whole articulation process.

2. METHODS

2.1. Participants

Ultrasound data were collected from three Standard Mandarin speakers (2 female, 1 male, ages 21-23) and three Hangzhou Mandarin speakers (2 female, 1 male, ages 21-23). The control subjects were from Beijing and speak only Standard Mandarin without any dialects or varieties of Mandarin. Native Hangzhou Mandarin speakers were born and raised in Hangzhou city and had never left the Wu dialect-speaking region for more than one year. None of the participants reported any speech deficits or auditory impairments.

Vowel	/n/ coda		/ŋ/ coda	
	Monosyllabic	Disyllabic	Monosyllabic	Disyllabic
/i/	因 **in**55	会心 xweɪ51.**ɕin**55	英 **iŋ**55	彗星 xweɪ51.**ɕiŋ**55
/ə/	根 **kən**55	出身 tʂʰu55.**ʂən**55	耕 **kəŋ**55	出生 tʂʰu55.**ʂəŋ**55
/ia/	烟 **ian**55	走线 tsoʊ214.**ɕian**51	央 **iaŋ**55	走向 tsoʊ214.**ɕiaŋ**51
/a/	安 **an**55	一般 i51.**pan**55	肮 **aŋ**35	一帮 i55.**paŋ**55

Table 1: Stimuli.

2.2. Materials

The stimuli consisted of 4 pairs of monosyllabic words and 4 pairs of disyllabic words (with target nasal coda in the second syllable), differing only in their nasal codas. The target syllables were in VN structure or started with a plosive or fricative consonant. To prevent the impact of assimilation, no other nasal sounds appeared in the target words (shown in Table 1).

A carrier sentence was used to present the stimuli in simplified Chinese characters without Pinyin or IPA annotation: __, 我知道怎么读__. ([uɔ ʈʂi.taʊ tsən.mɤ tu] "I know how to pronounce"). The printed-out reading material presented all these sentences in randomised order. For the sake of the participants' comfort, the recording session of each speaker was controlled to around 40 minutes, which only allowed speakers to repeat the material twice. Therefore, there are 384 tokens in total (16 target words * 6 speakers * 2 sentence positions * 2 repetitions).

2.3. Procedures

The ultrasound experiment was carried out in a sound-attenuated booth of the Fudan University Phonetics Laboratory in Shanghai. Ultrasound images and audio were recorded simultaneously using the Articulate Assistant Advanced system [18]. A Telemed MicrUs portable ultrasound device outfitted with Telemed MC4-2R20S-3 convex probe was used to collect midsagittal ultrasound data with a 90° field of view at a rate of 90 frames per second. Articulate Instruments headsets were used for probe stabilisation. An AKG C544 microphone was mounted on the headset to the right of the speaker's mouth to record 16-bit audio at a sampling rate 44.1kHz.

Before the experiment started, each participant had 30 minutes to familiarise the reading material and adjust their sitting position. The reading list was printed on one page to be conveniently held by subjects. An operator controlled the ultrasound and audio recording.

2.4. Data analysis

The annotation of audio files was made by hand in Praat [1] and the segmentation information was saved to Textgrid scripts. Considering the difficulty of dividing nuclear vowels and the following nasal codas in Mandarin, the target segments were set to include both vowels and syllable-final nasals. The feasibility of this segmentation criterion lies in the fact that each pair of target syllables differ only by their coda nasals, so the distinction of tongue movement in the articulation only comes from the different places of articulation of their coda nasals.

Ultrasound recordings were cut into segments based on the corresponding Textgrid script. Then, these ultrasound video clips were converted to a sequence of continuous images.

For data processing, these ultrasound images were imported into Python [16] as 8-bit numeric matrices. All pixel coordinate information and intensity values from each image were collected for further analysis. The tongue surface in a midsagittal ultrasound image always appears as a bright curve, which is manifested as a group of pixels with greater intensity value than its surrounding in the matrices. Suppose we draw a one-pixel width line crossing roughly perpendicular to the tongue surface (as the white line in Fig.1b) and extract all the intensity values of pixels on the line. In that case, the peak of this group of values indicates where the tongue surface is. This line is referred to as the line of interest in the following.

The alveolar and velar nasals are articulated by lifting parts of the tongue body to touch the alveolar ridge or soft palate, respectively. The tongue undergoes an upwards movement. We mark the bottom point of the line of interest as point A (see Fig.1, b). In this paper, we refer to the distance from point A to the intersection point of the tongue surface and the line of interest (shown as the coloured points

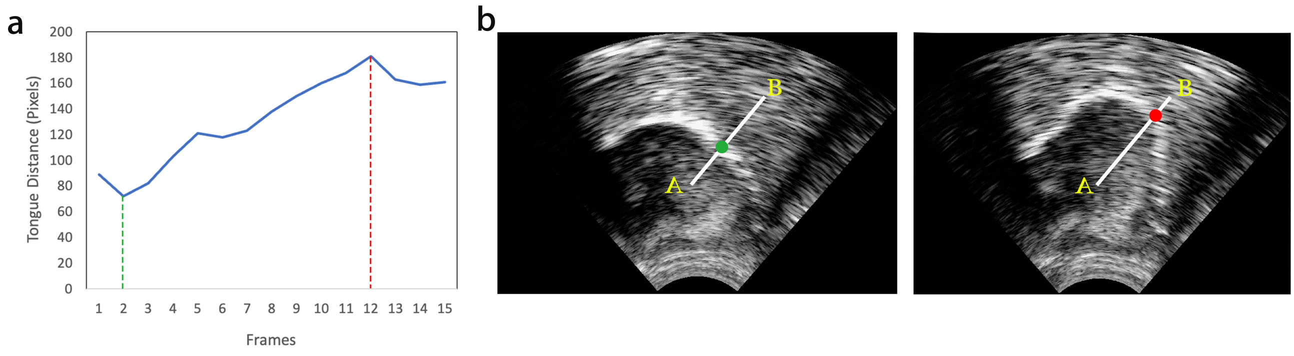

Figure 1: (a) The tongue distance varying over frames on a time series. (b) The line of interest for analysis is denoted between points A and B. The image on the left corresponds to the moment of minimum tongue distance. Likewise, the image on the right corresponds to the moment of maximum tongue distance. The tongue tip is on the left.

in Fig.1, b) as "tongue distance", which indicates the relative position of the tongue surface to point A. The greater the tongue distance, the further the tongue surface is to point A. Therefore, the change of tongue distance over time presents the tongue movement in the direction of the line of interest.

By calculating this tongue distance metric for each ultrasound frame in the target segment, we obtain a dynamic metric of tongue position that demonstrates movement over time (Fig.1, a), which we refer to as the tongue movement trajectory.

The line of interest should be set to explain specific research questions. In this experiment, we are interested in the tongue movement to the alveolar and velar nasal places of articulation. For every participant in this experiment, there are two lines of interest, pointing at the alveolar ridge and soft palate, respectively. Thus, for each participant, we generate two tongue movement trajectories: one for alveolar articulation and one for velar articulation. And to fully track the tongue movement, the line of interest should cover each speaker's whole range of articulation. Taking all the conditions into account, the length of all the lines of interest in this experiment is fixed at 200 pixels while the exact angle and location are adjusted to the physiological characteristics of individuals.

After measuring the movement trajectory data of every token, data cleansing is conducted by filtering out the outliers that have values lesser/greater than 1.5 times the interquartile range. Time normalisation is subsequently applied to all tokens by extracting the tongue distance at decile time points of the target segment. To remove the physical variation among all individuals, the tongue distance data are normalised using max-min normalisation. Each speaker's minimum and maximum is determined by the range of tongue movement throughout their entire data set.

For each speaker, local regression (LOESS smoothing) is performed on the normalised distance data grouped by alveolar and velar nasal coda to display the tongue movement when articulating these two nasal codas.

3. RESULTS

Fig. 2 shows the local regression results of the tongue movement trajectories in alveolar (a) and velar (b) places of articulation for every speaker. The colour-filling area represents the 95% Bayesian confidence interval.

As seen in Fig. 2, the alveolar and velar nasal coda tongue movement trajectories for Standard Mandarin speakers start from comparatively close proximity and go towards separation. And there is very little overlap between the alveolar and velar nasals trajectories, indicating that the tongue undergoes significantly different movements in articulating these two coda nasals. So in Standard Mandarin, the place of articulation of syllable-final /n/ and /ŋ/ are articulatorily distinct.

As a comparison, the two tongue movement trajectories of alveolar and velar coda nasals for Hangzhou Mandarin speakers overlap with each other throughout the articulatory segment, in both alveolar and velar places of articulation. It suggests that the Hangzhou Mandarin speakers display the same dynamic articulatory patterns for both syllable-final alveolar and velar nasals. There is no significant difference between the place of articulation of the alveolar and velar coda nasals in Hangzhou Mandarin, suggesting that the nasal coda is neutralised at both of these places of articulation and throughout the entire VN segments.

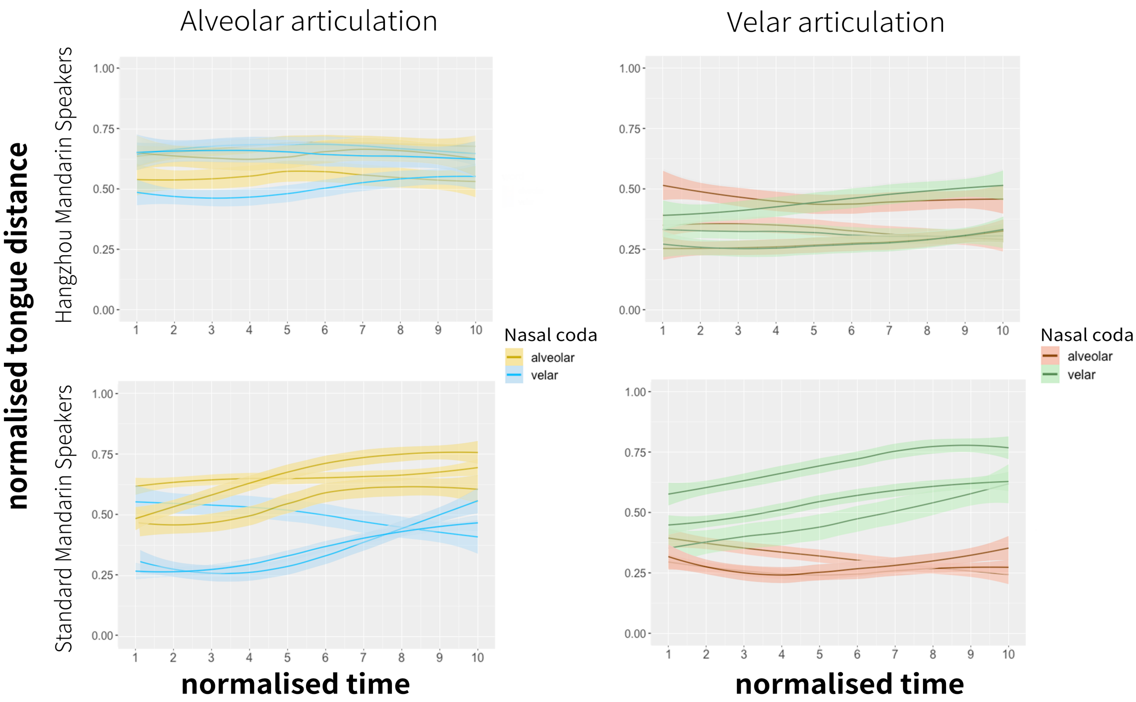

Figure 2: The regressed tongue movement trajectories of the nasal coda articulation for Standard Mandarin speakers (the bottom row) and Hangzhou Mandarin Speakers (the top row) in the place of articulation of alveolar (the left column) and velar (the right column). The x-axes are normalised time, and y-axes are normalised tongue distances.

4. CONCLUSION AND DISCUSSION

In conclusion, this study supports the existence of nasal coda mergers in Hangzhou Mandarin. Compared to other research, this dynamic study provides analysis reflecting tongue movement during the entire articulation process of the VN sequence, rather than the static tongue gesture on a single point.

However, the determination of the merger direction requires additional tongue contour analysis. Although this study does not address this issue, it can serve as a reference for future tongue contour analysis. The tongue movement trajectories provide an ideal method for determining representative frames for generating tongue contours; rather than choosing "magic moment" frames in an a priori fashion, they can be selected in a bottom-up data-driven manner. For example, in this case, to examine whether the tongue shapes of alveolar and velar nasals are merged in Hangzhou Mandarin, the most representative time point should be at the moment when the tongue movement trajectories of alveolar and velar nasals differ most. Images extracted from these time points could be used as representative frames in tongue contour analysis.

Besides, previous studies proposed that the proceeding vowel context may trigger this neutralisation [9,12,17,19]. This new approach allows us to explore this question further, and to give more precise descriptions of the articulation of nasal coda in various vowel contexts.

The method itself also has much room for improvement. For example, in this experiment, the determination of the tongue surface position along the line of interest is merely through finding out the maximum pixel values. However, this pixel value maximum can be influenced by image noise. Further research could improve the stability of automatically identifying the tongue surface locations by investigating alternative methods.

5. REFERENCES

[1] Boersma, P. 2002. Praat, a system for doing phonetics by computer. Glot International, 5.

[2] Chao. Y. (1928). *Studies in the Modern Wu-dialects*. Beijing: Commercial Press.

[3] Chen, C. (1991). The tonal change and diffusion: From the sampling of different age groups in Taipei. *Journal of Chinese Language Teachers Association*,16,69-99.

[4] Chen, M. Y. (2000). Acoustic analysis of simple vowels preceding a nasal in Standard Chinese. *Journal of Phonetics*, 28(1), 43-67.

[5] Chiu, C., Lu, Y., Weng, Y., Jin, S., Weng, W., & Yang, T. (2019). Uncovering syllable-final nasal merging in Taiwan Mandarin: Ultrasonographic investigation of tongue postures and degrees of nasalisation. In Proceedings of *the 19th International Congress of Phonetic Sciences* (pp. 398-402).

[6] Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America*, 120(1), 407-415.

[7] Duanmu, S. (2007). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

[8] Faytak, M., Liu, S., & Sundara, M. (2020). Nasal coda neutralization in Shanghai Mandarin: Articulatory and perceptual evidence. *Laboratory Phonology*, 11(1).

[9] Guan, Y. (2019). Nasal coda realization in speech production of Shanghai Mandarin. In Proceedings of the *19th International Congress of Phonetic Sciences*, Melbourne (pp. 467–71).

[10] Hsu, H. (2007) Syllable-final nasal nergers in Taiwan Mandarin-leveled but puzzling. *Concentric: Studies in Linguistics*,33(1), 1–18.

[11] Li,M., Kambhamettu. C, & Stone.M. (2005). Automatic Contour Tracking in Ultrasound Images, *International Journal of Clinical Linguistics and Phonetics*, 19 (6-7), 545-554.

[12] Luo, M., Sloos, M., & van de Weijer, J. (2015). Perception and production of Mandarin nasal codas by Shanghainese speakers. Chinese Accents and Accented Chinese CAAC2014, 16.

[13] Mielke, J. (2015). An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *The Journal of the Acoustical Society of America*, 137(5), 2858-2869.

[14] Qian.N.(1992). *The Research of Contemporary Wu*. Shanghai: Shanghai Education Press

[15] Tse, J. P. (1992) Production and perception of syllable-final [n] and [ŋ] Mandarin Chinese: An experimental study. *Studies in English Literature and Linguistics*,18,143-156.

[16] Van Rossum, G. & Drake, F.L. (2009). Python 3 Reference Manual, Scotts Valley, CA: CreateSpace.

[17] Wang, L., Cui, J., & Chen, Y. (2018). Wuxi Speakers' Production and Perception of Coda Nasals in Mandarin. In INTERSPEECH (pp. 2559-2562).

[18] Wrench, A. 2008. *Articulate Assistant Advanced User Guide* version 2.05.

[19] You. R. (2011). Phonetic and Phonological Features of Hangzhou Dialect. *Bulletin of Chinese Linguistics*, 5(1), 129–144.

[20] Zhengzhang, Shangfang, and Wei Zheng. (2015). *Wu dialect*. Oxford: Oxford University Press.